DOCUMENT RESUME

ED 118 612 TM 005 100

TITLE Symposium: The Care and Feeding of Test Constructors.
PUB DATE [Oct 74]
NOTE 39p.; Symposium presented at the Annual Military Testing Conference (16th, Okalahoma City, Oklahoma, October 21-25, 1974)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage
DESCRIPTORS Military Personnel; *Military Training; Program Effectiveness; *Quality Control; Surveys; *Symposia; *Test Construction; *Testing Problems; Training Techniques

ABSTRACT

Symposium papers presented are: (1) The Perspective of a Testing Psychologist Who Conducts Orientation Programs for New Test Constructors by Walter Birdsall; (2) The Perspective of a Military Supervisor of Test Constructors Who Are Temporarily Brought in from the Field to Write Tests by David Vaughn; (3) The Perspective of a Civilian Test Psychologist Working with Test Constructors by Carl R. Smith; (4) The Perspective of a Non-Commissioned Officer Who Is Actually Constructing Tests by J. R. Harris; and (5) The Care and Feeding of Test Constructors by George D. Oetting. Participants in the symposium were in the Army, Navy, and Coast Guard. Common concerns were methods to best train temporary examination writers, need for quality control, and identification of effective training materials and strategies. Survey of Test Constructors is attached. (BJG)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED118612 ...

SECTION XIV

SYMPOSIUM:
THE CARE AND FEEDING
OF TEST CONSTRUCTORS

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 100

THE PERSPECTIVE OF A TESTING PSYCHOLOGIST WHO CONDUCTS ORIENTATION PROGRAMS FOR NEW TEST CONSTRUCTORS

WALTER BIRDSALL

The comments which follow applied to the Naval Examining Center when it was a separate entity. Some of the concepts and procedures still apply, others are undergoing modification, and still others fall under the category of "that was yesterday". Naval Education and Training Program Development Center is a young organization and an exciting one. For those of us who had long standing loyalties to our previous identities, we have amazed ourselves at the quickness we have blended and evolved a new allegiance.

A training program for examination writers just did not happen one day. Between the extremes of a four-month program to a three-day general orientation, many programs were tried -- many discarded. The Exam Center used E-7, E-8, and E-9 chiefs on a normal tour of shore duty as writers. We had about ninety such chiefs. Duty rotation dates were not fixed conveniently at a specified time. One month maybe no new cheifs reported in, the next month maybe six. Not all chiefs had the same length of shore duty. Annually, we would have between twenty-five to thirty new chiefs who would have to be trained. However, the advancement examination cycle is fixed. It begins one year prior to administration of the examinations. Each writer is responsible for developing an examination for each of the four paygrades within his rating. The development takes about five months. With two examination cycles a year, little time is left for training. When a new chief reported in, it was necessary to get him to produce as quickly as possible. A chief, well seasoned in examination writing may be transferred in the middle of a cycle and be replaced by an untrained man. Therefore, one factor dominated and controlled all other considerations. Timing. When to conduct a training program. The "when" would determine what to teach and to a large degree how it was to be taught.

Our solution was a departmental manual containing the procedural steps of examination development--a cookbook. Devoid of philosophical considerations and nicities, the manual states that first you do this, then you do that, etc. No attempt was made to include preferred style, techniques of writing rules of grammar. The intent of the manual was to familiarize the new writer with the steps necessary to accomplish his job. We found we had to expand one chapter beyond the intended scope--Statistics. This one chapter is a layman's approach to the broad body of statistics and also specifically a "how to" with those statistics which the writer must use in developing his product. Mrs. Elberta Soloman wrote this particular chapter which is a gem because it gets a difficult subject across without talking down to the reader. After a new chief reported in and was assigned to his division, he spent the first week reading and rereading this manual under the guidance of his division Education Specialist. From then on for the next two months, it became "hands on" training--implementing the manual, under very close supervision of the Education Specialist, to actually produce or finish producing an examination. Four times a year the chiefs who reported in during the previous quarter would meet every morning for a week with the department head and/or deputy to discuss any possible weaknesses in the manual and any problems in implementing the manual. As the time being aboard varied among the chiefs, the content of the meetings ranged from basic indoctrination to true critique. As there were no more than eight people at these meetings, the meetings were very informal and unstructured. An equally important purpose of these meetings was to reinforce the valuable contribution they were making. In the never ending production of examinations, the purpose of the examinations becomes elusive if not illusive. The examinations themselves become the purpose. The writer cannot lose sight of the true goal of properly manning the fleet with the best of the qualified. The task of examination development is a unique opportunity for the Exam Writer to make a significant and lasting contribution to the Navy of today and the Navy of the

future--the man selected for advancement to Petty Officer Third Class today could well be a Master Chief Petty Officer in the Year 2000. This was the challenge presented in being assigned to the Naval Examining Center.

To maintain a check on the effectiveness of our training and to reinforce the concept of the uniformity of our testing product, we had a Qaulity Control Board. This board was a staff entity divorced from the Development Department. One of the major functions of this board was to assess the product of the department in terms of the standards of the department. This last point is important. It would be easy for such a board to evolve its own standards which could be different from those of the producers. Therefore, such a board must be vigilant that the standards used in assessing are truly the same as those of the producers. Any other posture is chaos, I assure you. I've been there.

The Quality Control Board must work closely with the Development Department in order that dual, or the appearance of dual leadership be avoided. In reviewing examinations, a board can ascertain the effectiveness of training of each division, can spot weaknessess that should be addressed. A good board functions as the major source of the effectiveness of the training of writers.

There was a weakness in our program--a weakness shared by anyone involved in developing examinations. Lack of resource material. In all of the books devoted to educational and psychological testing, there is little space devoted to actual item writing. The work of Robert Ebel is yet to be surpassed. However, his dicta are designed for people who are training for, or in, the field of Education and Psychology. Gronlund's work is a good extension of Ebel's, but still does not meet the need of a source for the non-professional.

We had started to develop two basic reference-text books. One book was to cover the basics of item writing; the other the grammar of questions. We had found that books on grammar do not quite meet the writing needs imposed by interrogatory rather than declarative

style. However, the circumstances of consolidation have overtaken their completion.

As you can see, the program which we had was centrally controlled and centrally monitored.

Now we are in a new situation. Other products, along with and tantamount to the examination, are being produced. Whether to train for creating all products or to train for developing individual products is still to be determined. But we have an advantage over other new organizations in that we know that we must have a training program and we are aware of the problems that must be accomodated and/or solved.

The Perspective of a Military Supervisor of Test Constructors Who Are Temporarily Brought in from the Field to Write Tests

1LT DAVID VAUGHN

A typical test construction project at the USAF Occupational Measurement Center lasts about six weeks, and five tests are written for a particular Air Force specialty. These five tests include four Specialty Knowledge tests: two tests for promotion to E-5, one for active-duty personnel and another for Air National Guard/Air Force Reserve personnel and two tests for promotion to both E-6 and E-7, one for active-duty personnel and another for Guard/Reserve personnel. The fifth test is an Apprentice Knowledge Test. One use for this test is for assessing knowledge of active-duty personnel if they have experience in civilian jobs which are related to their military specialties. Approximately five subject-matter specialists (SMSs) are brought to the Center on temporary duty from throughout the Air Force for the duration of the project. These SMSs, who are senior noncommissioned officers (NCOs) and who are experienced in the specialty for which tests are to be written, provide technical knowledge in the subject matter of the specialty. A test psychologist, who is permanently assigned to the Center, works directly with the SMSs to make up a test construction team. These test psychologists, who are junior Air Force officers, like myself, or civilians, provide knowledge and experience in test construction policies and procedures. During the course of a test construction project, the expertise of a test psychologist and that of the SMSs are combined to produce the tests.

Every two test construction teams are assigned a review psychologist, also military or civilian, who provides a first-line quality control check. All test outlines, items, master test copies, and other material produced by a test construction team are checked by the review psychologist. A senior review psychologist, who is responsible for monitoring a large number of career fields provides a second quality control check. The key to good test development under this system is teamwork among the SMSs, the test psychologist, the review psychologist, and the senior review psychologist.

During the course of a test construction project, a number of tasks are performed. Current editions of the tests are reviewed, Air Force specialty descriptions are reviewed, and outlines for the new tests are developed. Old items are reviewed and revised when appropriate, new items are written, items are selected to appear on each test, and the completed tests are reviewed and finalized. With all these tasks to be completed during a six-week period, there is little time for formal training of SMSs in item writing. Some sort of training at the Center is necessary, however, since SMSs usually have little or no previous experience in test construction. How is this training accomplished?

Before SMSs arrive at the Center, they are mailed a packet of information concerning their temporary duty assignment. Included is

a copy of the Handbook for Subject-Matter Specialists, which is published by the Center. The Handbook gives an overview of a test construction project. Included is information on developing the test outlines, old test items selecting and revising writing, new test items developing illustrated items, and reviewing and finalizing the test products. This Handbook serves as a useful reference for SMSs during a project, as well as helping them familiarize themselves with the test construction process before they arrive at the Center.

When SMSs first arrive at the Center, they are given a formal briefing. In this briefing, the Center missions are explained. The Weighted Airman Promotion System and the Weighted Airman Screening System are described and the relationship of the tests to these promotion systems is clarified. A brief overview of a test construction project is given in which the various phases of a project are outlined.

The Handbook and the formal Center briefing only serve as an introduction. The main responsibility for training of SMSs lies with the test psychologist. Although test psychologists have freedom in choosing the exact training procedures to be used, as a practical matter, most test psychologists use similar procedures. Most training is accomplished by short, informal briefings at appropriate points in the project, accompanied by extensive on-the-job training.

In the initial briefing given by a test psychologist, immediately after the SMSs arrive in the team room, SMSs learn what to expect during the project. The importance of the tests to the Weighted Airman Promotion System and the Weighted Airman Screening System is explained again. The responsibility of the SMSs for the final test products and the large impact that the tests will have on the future shape of the specialty is emphasized. SMSs are encouraged to always understand why they are doing something; since this will help them to do it better.

The first step of a test construction project is to review the current editions of the tests to be revised. The SMSs take the tests, and then go over each item with the test psychologist. The main purpose for this review is to delete obsolete or otherwise faulty items from scoring. However, this review is also an important training process. An SMS is able to get an approximate measure of his job knowledge and to gain insight into the actual experience of taking one of these tests from the standpoint of an examinee. An SMS is able to get a "feel" for what the tests are like, both in terms of format and content. At this point in the project, consideration can be given to what general sorts of changes should be made in the tests.

The next step involves development of a test outline and weighting of the outline areas to reflect the number of test items to be written on each outline topic. SMSs are briefed on the form of the outline and various guidelines in outline development, including a review of the specialty description. As the SMSs develop the outline, the test

psychologist is present to offer suggestions and to help resolve conflicts. At this point, consideration is given to any recent or potential major changes in the structure of a specialty and to what reference material is available for use. The review psychologist and senior review psychologist may also discuss these matters with the team.

After the outline has been developed, items on the current version of the tests are reviewed for reuse. These items may either be reused exactly as they appear on the current tests, in which case they are considered verbatim items, or the items, may be reused with variations in the wording, in which case they are considered revised items. The training pattern of informal briefing and extensive on-the-job training is very clear during this part of the test construction process. SMSs are briefed on the various factors which should influence the decision to reuse and/or revise an item. Interpretation of the item statistics and test statistics is very important. For example, after I have explained the item statistics, I put some examples from the tests to be revised on my chalkboard and help the SMSs interpret the data to decide if the items should be reused, whether they should be revised, and how they should be revised. Processing procedures for old items are also explained. SMSs are then given all the old items to review. As soon as the decision to reuse an old item has been made and a revision, if necessary, has been written, the test psychologist looks at the item and provides immediate feedback to the team concerning any errors or good characteristics in their work. I think that the main training occurs during these discussions of items between the test psychologist and the SMSs.

The next step is the writing of new items. Some test psychologists present briefings on principles of new item construction together with the briefing on old item selection and revision. Although this is a reasonable approach, since old item revision requires knowledge of item-writing principles, I usually wait until the old items have been processed before I talk formally about item writing. In this new item-writing briefing, various rules and guidelines of item-writing are presented along with examples. The Center has published the SKT Workbook for Subject-Matter Specialists as an aid in teaching the principles. This publication contains items which illustrate various problems, along with a discussion of the problems in each item and a better version of each item, where appropriate. To illustrate cluing of the correct answer in test items, the Workbook contains an examination in Franzipanics. Franzipanics is a subject that nobody knows anything about, but all the items can be answered correctly from hints in the wording and format of the items. For example, one item reads:

What is the primary purpose of the cluss in frumpaling?
(a) To remove cluss-prangs
(b) To patch tremalls
(c) To loosen cloughs
(d) To repair plumots

That answer (a), To remove cluss-prangs, is correct can be seen because it repeats cluss, a key word in the stem. Also, the correct answer is longer than the other choices and is the only choice that contains a hyphenated phrase.

Several test psychologists have prepared visual aids to help remind SMSs of the various item-writing guidelines while they are writing. Again, there is great emphasis on providing rapid feedback to the SMSs concerning their items. As items are written, the test psychologist reviews them and notes various problems and possible suggested changes. He then discusses certain items with the entire group of SMSs so that they may see, through practical example, what common mistakes to avoid and how to avoid them. Also, good characteristics of items are pointed out so that they may serve as models for other items. Various alternative item formats, particularly formats for illustrated items are discussed.

The same procedures for developing an outline, selecting old items, and writing new items are followed for each of the five tests. Although briefings are not usually given after the first set, the on-the-job training process of reviewing each item and discussing it with the item writer is followed throughout the entire project.

How well does this informal training procedure work? Can NCOs who are inexperienced in item writing be trained to produce a good test and, in addition, to produce five tests in six weeks? Yes, I think so. Of course, there is wide variation in how well a team learns to write items. I have had teams where after the first test, the changes that I made in items involved little more than corrections in spelling and grammar. On the other hand, I have had teams where I practically rewrote every item and, of course, had my revisions examined and approved by the SMSs. In all cases, I believe that our tests were of high quality and that they benefited from having been constructed by men who are fresh from the field and who know from first hand experience what their Air Force specialty is like today.

THE PERSPECTIVE OF A CIVILIAN TEST PSYCHOLOGIST WORKING WITH TEST CONSTRUCTORS

Carl R. Smith

US Army Enlisted Evaluation Center

Since the beginning of the Army's Enlisted Evaluation System in March 1958, the uses of evaluation results have constantly increased. In the beginning, the results were used solely to identify those soldiers eligible for proficiency pay. Since that time, promotion qualification, retention and reenlistment eligibility, special training eligibility, and other determinations have been based, in part, on evaluation scores. Today, the enlisted soldier's career largely depends on his performance as measured by various evaluation instruments. The written military occupational specialty (MOS) evaluation test is one of the main instruments used to evaluate soldiers.

Civilian psychologists in the Test Branch of the Enlisted Evaluation Center serve as principal contacts for MOS Evaluation Test Project Directors and their staffs at Item Writing Agencies (IWA) concerning matters relating to our Center and the Enlisted Evaluation System. We provide daily guidance via the AUTOVON telephone system and by mailings of FOUO and/or lengthy written information and guidance.

The enlisted evaluation written test cycle begins 14 months before the test administration month. This cycle begins with a form letter (subject: MOS Test Plans) sent from our Center to IWAs requesting test plans be submitted in accordance with a listing of MOSCs and a schedule of suspense dates for test plans in Inclosure 1. We discuss information requested in Inclosure 2 with the IWA test project director which includes proposed or recommended changes to MOS structures, enlisted soldier strengths in MOSCs, job descriptions, equipment items, status of Army publications (DA and non-DA) under revision, and classification of tests when other than FOUO.

These test plans are a very important step in the test development cycle. The Headquarters, US Army Training and Doctrine Command (TRADOC) sets and directs policy for Army service schools. They have directed schools to "systems engineer" all MOSCs for which those schools hold proponency. Relating to test planning, a systems engineered MOSC should contain a list of all duty positions and tasks.

Further, all such positions and tasks should be referenced to DA directives. Item writers, under the direction of the test project director at each IWA, should use the systems engineered study as the "keystone" of their work. Other regulations or documents to review for test planning should include, but not be limited to, job descriptions in AR 611-201/611-202 and the Military Occupational Data Bank (MODB) which lists task statements of an MOSC.

Ten months before a test administration month, a second request letter with two inclosures is sent from USAEEC to the IWA. Inclosure 1 is the schedule for submitting study references and item banks to USAEEC for a specific evaluation month or period. Inclosure 2 gives instructions for use of scheduled MOS test actions. Other inclosures attached to this letter are the MOSC test plans for the IWA for this evaluation period.

We develop a study guide from the references provided by the IWA. References are furnished to us for each skill level in the MOS. We merge a test plan, less numbers of items for areas, with each skill level (MOSC) reference list to make up this study guide. Also, certain standardized information about the purpose, scope, study material, study habits, training courses, score reporting, and correspondence study courses is included in this study guide.

Five fundamentals of item writer training which should be accomplished during FONECONs or TDYs with IWA test project directors and item writers should include:

1. Establish a good working relationship by being friendly, sincere, and enthusiastic about developing good quality test materials. Encourage item writers to do their best in learning their jobs and in developing their item writing abilities as rapidly as possible.

2. Set standards of performance which are fair and logical and will challenge their skills.

3. Frequently encourage item writers to demonstrate their skills within their capabilities.

4. Check as often as necessary to show that we (TPD and psychologist) are interested in the item writer's progress.

5. During TDY, give the item writers on-the-job training (special coaching) which may be required to meet the understood standards of performance.

The USAEEC psychologists take two TDY trips annually to each IWA. During these trips, they talk with the item writers, especially new item writers, to find out what is bothering them about the job, and

then check these findings with their supervisor and/or test project director in the MOS test branch. Those training needs that fall within our responsibility should be noted and a brief training program developed to satisfy those needs. The psychologists conduct as much of this training as possible during the TDY week. If unable to conduct all the needed training during the working time available, assign the new item writers to the senior item writers who will serve as instructors.

Giving Instructions

Inform item writer instructors about the necessity to carry out accurate instructions to item writers as fast as they can understand them. If in doubt whether these instructions are understood, question the item writer. The minutes lost in making certain that instructions are understood may be regained many times, for then the item writers' mistakes will not have to be corrected at a later time.

Quality Guidelines

1. Set high quality standards and make certain that item writers understand them.

2. Get to know the writers. Advise them on how they can improve and give them recognition when they do.

3. Don't take less than the best. If an IWA sends in test materials that are not up to par, don't let it pass. Find out why. Maybe the item writer needs training or motivation is lacking. Friendly counsel may change the item writer's attitude. When the cause is found, some corrective action can be applied to it.

4. Keep communications moving. Give examples of the consequences of incorrect or incomplete communications, resulting in poor quality work. Help the item writers to recognize good quality by using specific examples.

5. Thoroughly inspect all materials submitted by the IWA to insure good quality and at the same time check the writer's performance.

6. Continually stress the necessity of an effective quality control program. When an item writer's idea is used, be certain to give the credit due him or her. Also, letters of appreciation should be written and presented to outstanding item writers by a responsible senior officer at IWA.

7. A good quality control program links every senior officer, every supervisor, and every item writer at the IWA. The links of such

a chain are strengthened by cooperation and coordination of efforts toward good quality test materials.

8. Maintaining good quality requires consistency in supervision. Psychologists, and the TPD, should assign item writers workloads that keep them busy. If the work assignment is too easy, the item writer develops an attitude of indifference.

9. Let the item writers know that they are the subject-matter experts, that we depend upon them, that we want to keep them informed, and that we want information from them that will improve test materials.

10. Item writers want to think that and to be told that they are doing a worthwhile job. The quickest way to kill initiative is to make them think that their work doesn't matter. The alert psychologist stresses the value of every job of every item writer that he or she works with.

11. Item writers like to work with a psychologist and TPD who assigns a job, trains them on how to do it, and then lets them go ahead on their own.

Guidelines to Building a Good Working Relationship Through Communications with IWA

1. Don't start talking before you know what you are going to say. Know what you want and plan the sequence of steps necessary to attain it. Study the assignment until it is so clear in your mind that you have no difficulty explaining it.

2. Identify your communications objective. Before you begin to communicate, ask yourself what you want to accomplish. Are you seeking facts? If so, decide what information you require so that you can ask intelligent, precise questions. Do you want to make a special assignment? Make certain you have analyzed the job thoroughly so you can explain it properly.

3. Don't do all the talking. Give other people a chance to tell what is on their mind. The only way you can convince people of your interest in them and your respect for their views is to hear their opinions and ideas. The psychologists who pay attention to what IWA people are saying learn more about them and more about their jobs. Remember that the purpose of communications is to build a foundation for "doing."

THE PERSPECTIVE OF A NON-COMMISSIONED OFFICER WHO IS ACTUALLY CONSTRUCTING TESTS:

ADCM J.R. HARRIS
INSTRUCTOR
U.S. COAST GUARD INSTITUTE
OKLAHOMA CITY, OKLA.

Before beginning my discussion, I would like to briefly describe my background. I have been in the Coast Guard for nearly 20 years. My specialty is aircraft maintenance, and for the last several years before coming to the Institute, I have been an aircraft maintenance supervisor. My duty tour at the Institute as a permanent duty instructor will be from 3 to 4 years.

My objective is to describe the problems that I have had as a new item writer. I will briefly describe our internal organization so that you can see where I fit into it and hopefully get a better "feel" for the problems I will mention.

Since the Institute is responsible for correspondence course development and servicewide exams, my time is divided between the two. I am required to prepare the E4 through E6 exams during July and August of each year and the E4 through E9 exams during February, March and April of each year. The rest of the year I develop correspondence training courses for my rating.

I begin an exam series by developing my individual exam outlines using as a guide the Coast Guard Enlisted Ratings Qualifications Manual that lists the skills required of each rating and pay grade. A warrant officer reviews the exam outlines for technical content and correctness. With the assistance of the exam evaluation staff and the test psychologist, I review the exam outline and statistical data on the previous exam to see where adjustments should be made in the revised exam. Final branch approvals are made by a middle grade line officer assigned as branch chief to each of the five branches - Deck, Engineering, Electronics, Clerical, and Aviation. When an exam is completed, it is reviewed by a GS-7 or 9 training specialist, who has a degree in English, journalism, or education. The completed exam then follows the same reviewing process as did the outline.

The adjustment problems for a new item writer vary from instructor to instructor, but the ones I have had are typical, I believe. One of the most difficult problems was being able to write questions with clear stems and plausible distractors. I knew basically what I wanted to ask, but I sometimes had difficulty translating that idea into a good question with plausible distractors. It has been my experience that the instructors who have the least amount of trouble in this area

are the ones who come to the Institute with at least a fair background in basic English grammar. Our unit has a crash course in English grammar for new instructors to help them overcome this problem area.

When a new instructor reports aboard, he is dazzled by new terms and computer readouts that seem unfathomable at first. Such terms as skewness, kurtosis, P-value, discrimination, and reliability are confusing, to say the least, to a man who was supervising aircraft maintenance only a few weeks earlier. In my own case I developed about two series of exams before I felt that I really understood what the statistical data meant and how I could get the most out of it. I develop a series every 6 months, so I am speaking of one year of experience and approximately 1000 items. The Institute gives a very brief introduction to statistics to new item writers and an instruction booklet for reference. Actually using the statistics seemed to be the best teacher for me.

Another problem for a new instructor is determining the weight or the number of items to ask on each subject area of an exam. Rates in the Coast Guard are not nearly as specialized as they are in the larger armed forces. The wide range of skills covered by each rating can cause some men who are in the same rating but have differing assignment patterns to become more proficient in one aspect of their rating than in another. I have to guard against overweighting the sections in which I have had the most experience or the parts of my rate that appeal to me the most. In some skill areas, it is more difficult to write good questions, and a natural tendency is to shy away from these areas.

Creating a test item for a specific pay level gave me some difficulty and also some surprises when I first began. I was amazed at how some items performed. Some items that I thought should be easy for a particular pay grade turned out to be difficult, and vice-versa. I found that generally in my rating the lower pay grades seemed to do better on theory questions and the upper pay grades seemed to do better on the practical questions. These things can be learned only through experience as an item writer, since each rating has different requirements.

Keeping current on changes in the field can be a real problem for permanent duty instructors. In some ratings, such as electronics and aviation, changes occur at a rapid rate. It is very easy for the instructor who has been out of the field for a while to test on obsolete information. About the only correction factor we have for this is an occasional field trip or training school. Beyond that it is up to the instructor to try to keep up with field changes.

The fact that most of the specialties have only one representative at the Institute can sometimes be a problem. Since there is no one else in his specialty assigned, an item writer is forced to go with his own

opinion at times. An inherent problem with this is the possibility of tunnel vision or "not being able to see the forest for the trees."

Another closely related problem is that since there is generally only one man per rating, quite often a transferring instructor departs before his relief is on board. The new man has absolutely no transition time to learn the new job. Since development schedules for examinations must be met, the man must sometimes do the best job he can with no training at all. Quite often an instructor must complete his predecessor's courses or exams without the benefit of notes outlining his predecessor's intentions or ideas.

A problem that affected the quantity of items produced was a lack of typing ability. Before coming to the Institute, I had taken a basic course in typing and had some typing experience. Although my typing speed has increased, at first it definitely limited the number of items I produced.

A major problem area for me as a new item writer did not relate to the mechanics of item writing but to an area that is generally ignored in discussing the problems of a new item writer - the problems of psychological adjustment. These problems, I believe, can and do have an impact on an item writer's job performance.

About all that the typical new item writer knows of the Institute before reporting aboard is that the correspondence study courses and servicewide exams for the Coast Guard are prepared there. I was somewhat apprehensive about going to a job for which I had not been trained and for which I might not have the aptitude to perform. I had had no formal schooling for quite a few years, with the exception of an occasional service training school. Therefore, there was some question in my mind as to whether or not I would be able to perform in a training environment.

I had a lot of difficulty adjusting to a desk job. Up to this point in my career, I had never had a job that required me to be at a desk for 8 hours a day. I think that most of the nonclerical rates find this to be somewhat of a problem.

Another problem that may seem rather trivial but is a real problem for some instructors is one relating to a loss of status. Since an item writer is a senior enlisted man, he has generally had subordinates working for him and quite often has been the top enlisted man in his particular work area. Because of the nature and the mission of a testing organization, nearly the entire work force consists of officers and senior enlisted men. A crew leader must convert back to a worker status, so to speak, at a point in his career and at an age that is not particularly conducive to changes of this kind.

A natural question to ask at this point would be, "In view of these problems mentioned, do our exams perform at an acceptable level?" This is a critical point since a man's standing on promotion eligibility lists is determined in large part by how well he does on the exam. Statistics provided by our test psychologist verify that the exams have performed acceptably. There were 92 different exams administered in our March 74 series. Forty-six of these had statistical populations less than 20, and 46 had populations of more than 20. Of the 46 with populations less than 20, 78% performed acceptably in the 3 areas of reliability, standard deviation, and mean. In the 46 exams with populations greater than 20, 95% showed statistical data above the acceptable level. Since statistical data is unreliable in the low population exams and half the exams are in this category, a great deal of reliance is placed on the instructor in our system. With an overall 87% exam efficiency, I would say that our instructors are overcoming most of their transition problems.

In conclusion, I have mentioned eleven problems that I consider the most difficult for a new item writer to overcome. Five of these problems are: (1) writing clear stems and plausible distractors, (2) understanding statistical data, (3) giving the appropriate weight to each exam section, (4) writing items for a specific pay level, and (5) keeping current with field changes. Two other problems relate to having only one instructor per rating: (1) "tunnel vision" and (2) no transition time for a new instructor. A lack of typing ability can be a problem for some.

Three of the problems I mentioned are psychological in nature. One of these is the apprehension involved in making a radical change to a job you are not sure you can perform. Also, there are the problems of adjusting to a desk job and the loss of status for some instructors.

Some of these problems can have a tremendous effect on the product the new item writer produces. Most of these problems are only temporary for most new item writers. By acknowledging that transition problems do exist for the new item writer, management can develop procedures and training programs that will help new men and will help in maintaining a more effective item writing staff.

"The Care and Feeding of Test Constructors"

By Lt Col George D. Oetting
Extension Course Institute
Gunter Air Force Station, Alabama

1. Introduction

Written tests permeate the military personnel system. From the time an individual takes tests to qualify for enlistment in the armed forces until he completes a full career, he is continually tested for classification, promotion and other purposes. Many of these tests are constructed by senior enlisted military personnel who know their jobs well and are highly proficient in the technical knowledge necessary to perform their operational duties. However, very few are professional experts in test construction procedures. Consequently, assignment to such duties may create special problems for newly assigned test constructors and the testing units to which they will be assigned.

These problems are the focus of this research. It is an initial attempt to explore this area within a number of testing activities of the armed forces. It includes the results of a telephone survey of testing units, a review of printed materials used to help orient new test constructors, and the detailed findings of a survey questionnaire of enlisted personnel assigned test construction duties.

2. Telephone Survey of Testing Units: To get an overall "feel" for the nature of testing programs operating in the military, a survey was conducted in March 1974 by telephoning major testing units in each branch of the military. Since military tests are used for many purposes, it was decided to concentrate on the following units which are deeply involved in tests used for promotion of enlisted personnel:

Military Branch	Unit Contacted
Army	Enlisted Evaluation Center
Navy	Naval Education & Training Program Development Center
Air Force	USAF Occupational Measurement Center
Marine Corps	Marine Corps Institute
Coast Guard	Coast Guard Institute

The following general areas were discussed with each unit:

a. Grade Level of Test Constructors: Most enlisted test constructors are in grade E-7, and above although others are occasionally used if necessary.

b. Assignment Status: All but the Air Force have test constructor personnel assigned on a stabilized tour of duty, lasting 2-4 years. Air Force personnel are brought in from operational units for 4-6 weeks temporary duty to write tests.

c. Use of Tests: Although these tests are used for promotion, their use in the promotion process varies. In some cases, the test score is the main factor in promotion; in others, test results are used as one factor in a promotion board and also for classification purposes.

d. Test Constructor Orientation: All units have developed an orientation program for newly assigned test constructors lasting 1-3 weeks. Quite often this is combined with on-going in-service programs throughout the tour of duty. In some cases, special workshops are conducted.

e. Test Design: All primarily use multiple choice items with 100-150 questions in each test. There are some variations, particularly in performance rating situations.

f. Testing Unit Organization: The overall promotion testing function is centralized in all branches of the military. However, those actually preparing the tests may be located at other units, such as in the Army; or the test constructors may be pulled in for testing duties, as done by the Air Force. All branches though, do have one unit designated as the final producer and controller of promotion tests.

3. A Review of Test Orientation Materials

There are many good reference books available in any academically respectable library on the principles and techniques of test construction. As part of this research, it was considered appropriate to take a look at the kinds of materials and activities used in the orientation programs of newly-assigned military test constructors. Therefore, the units contacted were asked to provide copies of the booklets, work sheets etc., that they used. Although bibliographic references were made to civilian texts, it was found that almost all material used was locally prepared. A review was made of materials received, and the following is a summary of the types of materials and orientation activities found in most units: (It should be emphasized that the summary cannot be considered comprehensive since it was not possible for the units to send all materials.)

a. Reference Materials Used:

1) Introductory information on testing philosophy and role of test constructor in the process.

2) General explanation of the overall sequence of local procedures to be used in test development.

3) General rules of good test construction with examples of poor and improved items.

4) Discussion with examples of the different types and formats of test items.

5) Suggestions on developing good distractors with examples of poor and improved versions.

6) Rules on use of illustrations and diagrams in tests with poor and improved examples.

7) General explanation of statistical concepts and terms used in evaluation programs.

8) Explanation with examples on the use of local test analysis statistics.

9) Explanation and examples of local administrative codings for test item cards, reports etc.

10) Procedures to be used in the safeguarding of classified and controlled test materials.

11) Check lists to be used as a final reference in the total sequence of test development.

b. Practice Activities:

1) Practice exercises in identifying errors in items and rewriting an improved version.

2) Practice exercises in identifying errors in either stem or distractors and rewriting an improved version.

3) Practice exercises in identifying and correcting errors in grammar, punctuation etc.

c. Particularly Interesting Orientation Activities:

The materials and activities listed above appear to be generally similar to the kinds of things mentioned in civilian texts. (With adaptations to local military requirements) This

is quite natural since the rules for good evaluation are the same regardless of the testing environment. There were some interesting activities however, a bit different from the usual, that are worth noting:

1) The "Murder Board" - Several units used a technique in which a group of test constructors get together to review all items proposed for a particular test. These are projected or visually displayed in some manner to allow all personnel to "murder" the item through a vigorous exchange of questions, comments and critiques. If the item survives this session (with any changes needed) then it is considered for inclusion in the new test.

2) Performance-Oriented Tests - Almost all units expressed concerns about the need for making test items as realistic as possible rather than a simple regurgitation of isolated bits of information. Several units have attempted to do this by developing several multiple-choice questions measuring the thought process, procedures etc. required to perform a particular challenging, realistic military task. Quite often something like a completed form or a case situation is presented beforehand and used as the basis for answering a series of questions.

3) Nonsense Tests - In teaching principles of test construction, some units presented test exercises with nonsense words in test items which nevertheless can be answered if a person understands what rules of test construction are being violated. Another unit requires test constructors to write items violating a principle, and then rewrite the item in a correct format.

4. Initial Thoughts on a Survey of Enlisted Test Constructors: Having attended five previous conferences of the Military Testing Association, this investigator has listened to many papers and discussions covering the whole spectrum of testing activities from predictive models to performance rating techniques. Much has been presented on the use, design and analysis of tests. Unfortunately, very little has been said about the people who actually write these tests. This may be due to the fact that most conference attendees appear to be in supervisory positions and are required to concentrate on the "big picture" testing problems. This gap in research is the focus of this survey. It is an attempt to get some personal feedback from the enlisted troops at the working level of test construction.

5. Survey Design:

a. The survey format was a combination of open-ended questions and a structured rating scale. It was printed on one page (both sides) and designed to be completed in 15-25 minutes. (See atch 1)

b. Survey questions were designed to secure the following information:

1) Prior Preparation: Prior personal contacts with enlisted personnel assigned test construction duties revealed that many had little prior formal training or experience in this area. This question attempted to see if this is typical of military personnel assigned to this work.

2) Test Constructor Orientation: Most military testing units have some sort of orientation program to help newly-assigned test constructors learn their duties. This forced-choice question required the identification of the activity considered most helpful to new personnel.

3) Toughest Current Task: As with any job, some tasks prove more difficult than others. This forced-choice question required the identification of the one considered most difficult.

4) Selection of Test Constructors: After personnel have had some test construction experience and worked with others in this process, they should have a pretty good idea of the prerequisites for a successful test constructor. This question requires an indication of the three most important selection critieria in priority #1, 2 and 3.

5) Difficulty of Test Construction Tasks: This question required subjects to take a look at the total spectrum of major test construction tasks, and indicate the degree of task difficulty on a nine degree scale. Additional space was provided if a subject felt any major task was omitted. This list was developed by reviewing civilian and military test references.

c. A preliminary draft copy of the survey form was sent to personnel in the major testing activities of all branches of the military for review and comments. Particular care was taken to insure that potential respondees were familiar with testing terminology used.

d. Since this was a survey of military personnel, authorization had to be secured from the various military branches to administer it. This was secured in May 1974 and the survey was assigned USAF SCN 74-136.

6. Description and Size of Population Surveyed:

a. The population surveyed consisted of experienced senior enlisted personnel who are responsible for writing tests to measure the technical knowledge of enlisted personnel in their own career fields. Each branch of the service has somewhat different organizational structures and purposes in its testing program; however, they all have enlisted personnel performing this function. The selection of the sample was determined after consulting supervisors in major testing activities of all branches of the military. The following units were sampled:

UNIT	NUMBER SURVEYED
Army	
Institute of Administration	12
Military Police School	11
Signal Center and School	9
Transportation School	9
Navy	
Naval Education and Training Development Center	98
Air Force	
USAF Occupational Measurement Center (Subjects were from various USAF units)	49
Marine Corps	
Marine Corps Institute	27
Coast Guard	
Coast Guard Institute	24
TOTAL	239

7. Administration of the Survey:

a. The survey was sent out to the units to be surveyed in June 1974. Specific subjects were not pre-identified; this selection was left to the units. A letter of instructions was sent with the surveys explaining the criteria and precedures to be followed. (See atch 2) A cover letter was attached to each survey, explaining the nature of the survey to the subjects. (See atch 3)

b. After the survey forms were completed, the units involved collected them and returned then to the Extension Course Institute for analysis.

8. Limitations of Research: As work progressed in this research, it became evident that any conclusions needed to be tempered with the realities of the military testing world. These limitations should be considered when reviewing the summary of results:

a. Sample Surveyed: Although a sample size of 239 could be considered statistically respectable, it may or may not be typical of the total population of military test constructors. Time and personnel restraints prevented a sampling of all activities using military test constructors. Those selected were thought to be typical of the major kinds of testing activities in each branch of the military.

b. Subjects in the Sample: The great variety in the organizational structures established to administer testing programs created an equally great variety in the ways subjects were involved in the test construction function. This ranged from full-time test construction duties to a one-time temporary duty of several weeks to write a test. Consequently, the responses were from people with varying degrees of involvement in this process.

c. Open-Ended Questions: The usual problems inherent in open-ended questions emerged when analysis was attempted. Some questions, (particularly #1 and 2) proved too broad and subject to varying interpretation; some responses were unusable because they related to local problems not in the context of this research. A total of the tabulations for questions #1-4 revealed that some subjects did not provide any responses to these questions. As a result any sophisticated analysis of differences between various groupings of subjects was not attempted.

9. Results of Survey: The results of this survey are presented below in tabular form for each question. These additional comments may help in interpreting the results:

a. Questions 1-4, being open-ended questions, resulted in many kinds of answers. These were grouped as much as possible into similar-type answers, and a frequency count made. Some subjects made no responses, or responses not pertinent to this research; consequently the frequency tabulations are not a 100% coverage of the subjects sampled.

b. In addition to the frequency tabulation for question #4, a weighted frequency column was added, since subjects were asked to list three tasks in order of priority. It is felt that those rated as priority #1 should be given a higher weight, to get a more refined indication of cruciality. Therefore a simple weighting system, explained at the bottom of the tabulation, was added.

c. The results of the task ratings in question #5 were presented in two ways to better depict the information. One indicates the percentage (of all responses to a task) which selected a particular rating block; the other indicates the mean rating (using a 1-9 scale) and the range of one standard deviation, above and below the mean. Almost all subjects completed all responses on this scale, so lack of responses was no problem with this question.

(TABLES START ON NEXT PAGE)

"1. Prior Preparation: How much background did you have in test construction prior to your current assignment? Please indicate this by using the following two blanks:"

"a. Formal training: (Colleges courses, Military schools etc.)"

TYPE OF PREPARATION	FREQUENCY
None	115
Military school assignment	81
College courses in this area	18
Bachelor Degree	3
College teaching experience	2
Associate Degree	1

"b. Experience: (prior instructor duties, test writing duties etc.)"

TYPE OF PREPARATION	FREQUENCY
Prior military instructor duty	102
None	84
Prepared locally - used tests	34
Wrote military courses	14
Prior participation in test revisions	6
Served as troop commander	2
Acted as test administrator	1

526

"2. Test Constructor Orientation: While you were going through the orientation program for new test constructors, there were various activities designed to develop your skills. What one did you find most useful?"

ACTIVITY	FREQUENCY MENTIONED
Using printed reference materials	19
Participating in lectures and discussions	17
Actually writing test items	14
Reviewing and practicing language skills	13
Interpreting statistical test data	13
Reviewing rules for writing test items	12
Preparing testing objectives and outline	11
Working with experienced test writers	9
Studying examples of good and bad items	8
Analyzing test items written	7
Constructing stem of test item	7
Writing test item at appropriate level	4
Taking old tests which are to be revised	2
Writing a variety of test item forms	1
Learning how to research materials	1
Writing task - oriented questions	1
Constructing plausible distractors	1
Learning how to use dictionary	1
Using common sense in writing items	1

"3. Toughest Current Task: As you now are constructing tests, what do you find to be the single most difficult task?"

TASK	FREQUENCY MENTIONED
Writing plausible, parallel distractors	62
Writing items to measure specified proficiency level in a career field	33
Researching reference materials for testable material	27
Writing a testing outline/plan	19
Insuring test items are current	16
Constructing new test items	8
Writing items to meet statistical requirements	7
Interpreting test analysis statistics	6
Preparing comprehensive tests	5
Using proper English and grammar skills	5
Writing the stem of an item	4
Writing an easy question	4
Devising items without use of schematics	4
Coding items for administrative purposes	3
Arranging items within a test	1
Using behavorial objectives to make a test	1
Using creativity in constructing items	1

"4. Selection of Test Constructors: If you were in charge of selecting future personnel to perform your work in test construction, what would be the three most important criteria you would use. Please list in order of priority - #1 being the most crucial."

CRITERIA	FREQUENCY MENTIONED	WEIGHTED FREQUENCY*
Broad experience and knowledge in career field	196	514
Positive attitude, desire for this type of work	104	210
Knowledge of English, grammar	35	73
Ability to write clearly and simply	43	72
Broad educational background	40	71
Prior instructor experience	34	67
Prior training in test item writing	28	53
Professional military bearing and conduct	16	30
High military performance reports	13	26
Creativity, imagination, flexibility	17	25
Senior in rank - E-9 if possible	13	22
12-18 years experience - not too old	11	19
Common sense - objectivity	9	14
Ability to get along with others	8	11
Availability and retainability for work	8	11
Administrative experience and ability	5	11
Good reading comprehension	4	10
Up to date in own career field	5	5
Pass a personal interview	2	5
Able to take criticism	4	4
Ability to type	4	4
Has no personal problems	4	4
Able to understand statistical analyses	1	2

*Weighted Frequency determined by multiplying priority by a weight factor

If priority is:		Weight Factor
#1		3
#2	Multiply by	2
#3		1

"5. Difficulty of Test Construction Tasks: In your current assignment, you soon discovered there are many different tasks involved in test construction. We have tried to identify the major ones in the list below. We would like your opinion on the degree of difficulty in performing each of these tasks, once you really got started in constructing tests. Please indcate this through a check for each task. (There are 3 blocks in each major division so you have a 9 degree scale available for each response)"

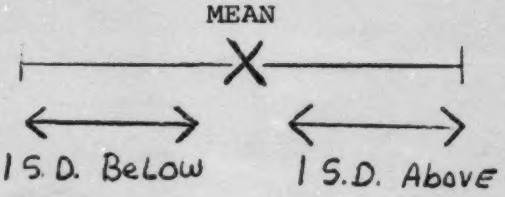
Test Construction Tasks	Degree of Task Difficulty									Not applicable to me
	Easily done -No problem			Average difficulty			Hard to do			
Creating a master plan (overall guide) for entire test.	7	11	10	(24)	20	11	9	1	2	5
Interpreting and using the master plan (overall guide).	13	14	(23)	18	16	6	6	1	1	2
Determining weights/coverage for different areas to be tested	8	10	14	(22)	15	13	10	3	0	5
Researching and verifying references used for each test item.	10	8	(18)	17	12	16	10	4	4	1
Understanding principles & statistics used in test analysis	6	4	10	(20)	16	12	17	6	5	4
Interpreting & using test analysis data to improve test items.	4	6	14	(18)	16	16	10	6	5	5
Preparing a clear stem for the item.	5	6	13	(29)	17	14	10	3	1	2
Preparing plausible distractors (foils).	4	3	7	15	10	16	(23)	9	10	3
Cross-checking relation of stem to all alternatives.	5	7	16	21	(24)	13	8	3	1	2
Applying general rules for constructing test items.	7	9	20	(25)	19	8	7	2	0	3
Editing, rewriting items already in use.	12	11	(24)	22	16	8	5	1	0	1
Creating a test item to measure a specified proficency level.	5	6	6	18	(19)	11	17	7	6	5
Selecting, designing diagrams, sketches etc. for use with items.	13	12	(19)	18	12	6	7	2	1	10
Final reviewing and assembling of all items into complete test.	17	12	18	(24)	14	6	3	0	0	6
Preparing student study guide related to areas tested.	9	6	13	8	10	4	5	1	2	(42)

The numbers above indicate the percentage of total responses which selected a particular block on the rating scale. The highest percentage for each task has been circled.

"5. Difficulty of Test Construction Tasks: In your current assignment, you soon discovered there are many different tasks involved in test construction. We have tried to identify the major ones in the list below. We would like your opinion on the degree of difficulty in performing each of these tasks, once you really got started in constructing tests. Please indcate this through a check for each task. (There are 3 blocks in each major division so you have a 9 degree scale available for each response)"

Test Construction Tasks	Degree of Task Difficulty: Easily done -No problem	Average difficulty	Hard to do	Not applicable to me
Creating a master plan (overall guide) for entire test.				
Interpreting and using the master plan (overall guide).				
Determining weights/coverage for different areas to be tested				
Researching and verifying references used for each test item.				
Understanding principles & statistics used in test analysis				
Interpreting & using test analysis data to improve test items.				
Preparing a clear stem for the item.				
Preparing plausible distractors (foils).				
Cross-checking relation of stem to all alternatives.				
Applying general rules for constructing test items.				
Editing, rewriting items already in use.				
Creating a test item to measure a specified proficency level.				
Selecting, designing diagrams, sketches etc. for use with items.				
Final reviewing and assembling of all items into complete test.				
Preparing student study guide related to areas tested.				

The figures above depict the mean rating and standard deviation for each task:



"(If you feel there are additional major tasks not mentioned above, please write these in below and indicate degree of difficulty.)"

ADDITIONAL MAJOR TASKS	Degree of Task Difficulty		
	Easily done -No problem	Average difficulty	Hard to do
Preparing understandable training objectives			X
Relating test items to rating qualifications			
Meeting format differences between course writers			
Systems engineering			X
Measuring different enlisted grades within same skill level			
Formulating objectives		X	
Writing clear student instructions			
Overall management of testing program			
Typing material for review			XXX
Limited reference materials			XXX
Keeping updated on career information		XX	
Preparing item with limited references			
Coding and administrative procedures		X	XX
Ability to use reference publications		X	
Analysis of tests with small population			X

The additional major tasks listed above are those added by subjects, which were not preprinted on the task list. In terms of the total number of subjects, they are not considered to be significant. Most of those receiving more than one "X" came from the same unit and appear to be more of a local gripe about final administrative requirements.

10. Summary of Survey Results: Keeping the limitations in mind outlined above, the following is a brief, subjective summary of the findings:

a. The majority of subjects have had no formal training in the test construction area. Those that have, obtained it through prior experience as a military instructor. Some have had experience in writing locally used tests or through writing military courses.

b. Subjects had divergent opinions on what was the most useful activity in orientation programs. Printed materials prepared by the local unit proved most useful. Also rating high were participation in group discussions/lectures and actually writing test items. Many felt the review of needed language skills and the interpretation of statistical data were valuable.

c. By far, the toughest task was felt to be writing plausible, parallel distractors for multiple-choice items. Two other tasks rating high in difficulty were writing items to measure a specified proficiency level within a career field, and finding suitable references with testable material.

d. Subjects clearly indicated that broad experience and knowledge in a career field was the most important criteria in selecting future test constructors. A positive desire and attitude for this type of work was considered crucial. Communication skills and a good educational background were also rated fairly high.

e. In rating all tasks on a difficulty scale, most tasks were considered in the average difficulty range. There were gradations however, with preparing plausible distractors, measuring specific proficiency levels and understanding test analysis statistics being rated most difficult. The easier tasks were interpreting the testing plan, editing old items, and using diagrams and sketches with items. Most felt that the task list covered the important areas of their work.

11. Potential Uses of Survey Information: It is hoped that the results of this survey will prove useful to other personnel involved in test construction. These are some of the possibilities:

a. Review of Orientation Program: The results indicate there are differences in the perceived difficulty of the various tasks in test construction. A review of your program may be appropriate to see if there are sufficient orientation activities to insure proficiency in the more difficult tasks such as writing plausible distractors and writing items to measure different proficiency levels.

b. Development of Orientation Program: If you are starting or revising an orientation program, the task list may help to provide some structure in planning the necessary activities. Most respondees seemed to feel that this list covered the major tasks involved in their work.

c. Improvement of Orientation Critique: The kinds of questions asked may prove useful in any critique forms you use to evaluate the effectiveness of your own programs. These could be easily modified to meet local needs.

d. Selection of Military Test Constructors: The criteria recommended by the military test constructors themselves are probably a realistic, honest appraisal of the qualities needed for this type work. They should be considered as you review personnel records and interview potential assignees.

e. Basis for Further Research: This survey was planned as an initial exploration of this subject. As such, the research design was not very sophisticated in terms of statistical treatment. Conclusions drawn on the basis of percentages and response tabulation cannot be considered as "proof positive". However, it is a start, and might be useful in developing more definitive research questions in this area.

f. Stimulate Further Survey Research: It was abundantly evident while conducting this research, that a great deal of time, money and personnel are devoted to testing activities within the Department of Defense. Many of the same kinds of functions are performed by these units. The survey approach used in this research could be applied to many areas of mutual interest. The pooling of research, experience and techniques in research reports could provide a valuable reference for others involved in the same area. It could stimulate interest develop personal contacts and help prevent duplication of effort.

SURVEY OF TEST CONSTRUCTORS
(USAF SCN-74-136)

1. Prior Preparation: How much background did you have in test construction prior to your current assignment? Please indicate this by using the following two blanks:

a. Formal training: (College courses, Military schools etc.)

b. Experience: (prior instructor duties, test writing duties etc.)

2. Test Constructor Orientation: While you were going through the orientation program for new test constructors, there were various activities designed to develop your skills. What one did you find most useful?

3. Toughest Current Task: As you now are constructing tests, what do you find to be the single most difficult task?

4. Selection of Test Constructors: If you were in charge of selecting future personnel to perform your work in test construction, what would be the three most important criteria you would use. Please list in order of priority - #1 being the most crucial.

#1.

#2

#3

5. Difficulty of Test Construction Tasks: In your current assignment, you soon discovered there are many different tasks involved in test construction. We have tried to identify the major ones in the list below. We would like your opinion on the degree of difficulty in performing each of these tasks, once you really got started in constructing tests. Please indcate this through a check for each task. (There are 3 blocks in each major division so you have a 9 degree scale available for each response)

Test Construction Tasks	Degree of Task Difficulty									Not applicable to me
	Easily done -No problem			Average difficulty			Hard to do			
Creating a master plan (overall guide) for entire test.										
Interpreting and using the master plan (overall guide).										
Determining weights/coverage for different areas to be tested										
Researching and verifying references used for each test item.										
Understanding principles & statistics used in test analysis										
Interpreting & using test analysis data to improve test items.										
Preparing a clear stem for the item.										
Preparing plausible distractors (foils).										
Cross-checking relation of stem to all alternatives.										
Applying general rules for constructing test items.										
Editing, rewriting items already in use.										
Creating a test item to measure a specified proficency level.										
Selecting, designing diagrams, sketches etc. for use with items.										
Final reviewing and assembling of all items into complete test.										
Preparing student study guide related to areas tested.										

(If you feel there are additional major tasks not mentioned above, please write these in below and indicate degree of difficulty.)

DEPARTMENT OF THE AIR FORCE
EXTENSION COURSE INSTITUTE (AU)
GUNTER AIR FORCE BASE, ALABAMA 36118

SUBJECT: Survey of Test Constructors

TO:

1. We greatly appreciate your willingness to administer this survey to people in your unit. The results of this survey combined with a review of service-wide procedures in this area will hopefully result in a useful study for everyone concerned. A report of this research will be presented at the Military Testing Association Conference in October; you will also be sent a copy of the study about the same time.

2. In administering the survey please use the following guidelines:

a. It should be given only to senior enlisted military personnel whose duties include writing test items on a regular basis.

b. Try to select people who still have a pretty good idea of the initial problems facing people learning how to write good test questions.

c. For your unit, we would like to survey people if possible. A few extra copies are also included.

d. Please explain the purpose of the survey and assure the respondees that their answers will be kept confidential and combined with all others being surveyed. The results in no way will be related to any individual or to any testing unit.

3. The survey is a short one - it should only take 15-25 minutes to complete both sides of the one page. We would appreciate your conducting the survey as soon as possible and return the completed forms in the attached envelope by This will give us time to tally and analyze the results and prepare the report for the conference.

4. This survey has been reviewed and approved for use in all branches of the Military. (Survey control Number: USAF SCN-74-136).

5. Thanks for your help. If you have any questions give me a call on Autovon, 921-4391/4392.

GEORGE D. OETTING, Lt Col, USAF
Chief, Educational Systems Branch

1 Atch
Surveys - Copies.

DEPARTMENT OF THE AIR FORCE
EXTENSION COURSE INSTITUTE (AU)
GUNTER AIR FORCE BASE, ALABAMA 36118

SURVEY OF TEST CONSTRUCTORS

Would you please take a few minutes to answer some questions about your experiences while learning how to write good test questions. As you undoubtedly already know, it is a difficult job to create an effective test. Maybe you felt somewhat like I did when sitting down to write my first test. I had taken hundreds of tests over many years of schooling and even had some courses in tests and measurement. But these experiences really didn't help much when I shifted from the role of test-taker to test preparer.

This problem is the focus of the attached questionnaire. We are interested in getting reactions of military personnel like yourself who have been assigned the tough but very important task of creating tests to measure the technical knowledge of military personnel. We are planning to query people from the Army, Navy, Air Force, Marines and Coast Guard who are involved in this work. Your reactions will be combined with all others and presented in a special symposium at the 1974 meeting of the Military Testing Association, hosted by the Coast Guard in Oklahoma City. Copies of the research will be sent to all units involved in the survey in October of this year. If you want a personal copy please drop me a note and I'll send you one.

This questionnaire has been reviewed and approved for use in your branch of the military. (USAF SCN-74-136)

Thank you for your assistance

George D. Oetting

GEORGE D. OETTING, Lt Col, USAF
Chief, Educational Systems Branch